MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2017/2018

HPB3011 –BIOINFORMATICS ALGORITHMS II

(All sections / Groups)

04 June 2018
9:00 - 11:00 AM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use a calculator.

Question 1

a) In a typical microarray experiment design, sample pooling is one of the commonly used strategies to reduce experiment cost. What is the caveat of the sample pooling in microarray experiment? [1 mark]

b) What is the difference between agglomerative and divisive clustering? [2 marks]

c) You performed 20 DNA microarray experiments to compare gene expression profile before and after drug treatment for 10 subjects. You want to know if there is any difference in gene expression before and after drug treatment. What statistical test is suitable for hypothesis testing? Explain your answer briefly. [1 mark]

Control condition (A)	Treatment condition (B)
Subject 1 (before)	Subject 1 (after)
...	...
Subject A_N (before)	Subject B_N (after)

d) Why are log ratios of gene expression values often used rather than raw ratios when analyzing relative gene expression level? [2 marks]

e) You are supposed to perform RNA-seq analysis of the root cells of *Arabidopsis thaliana* before and after salt stress. In a typical experiment, 30 million reads are sufficient to detect moderately expressed genes. However, you want to detect lowly expressed genes as well. Should you change the sequencing coverage? Explain your answer briefly. [1 mark]

f) You want to study transcriptional profile of a mouse during embryo development using RNA-seq. Should you prefer single-end or paired-end RNA-seq? Explain your answer briefly. [1 mark]

g) Give TWO advantages of RNA-seq compared to DNA microarray for the study of transcriptomics. [2 marks]

Continued .....

Question 2

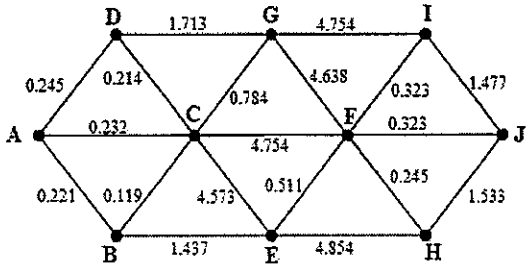
a) What are the two disadvantages of k-means clustering? [2 marks]

b) Given the following probabilities, use the Bayes approach to find the probability of having diabetes if the test is positive – P(Diabetes |Test+). [3 marks]

	Diabetes	NO Diabetes
Test +	0.8	0.01
Test -	0.2	0.99

Overall probability of diabetes = 0.01

c) The graph below illustrate the Euclidean distances among the genes (A,B,C,D,E,F,G,H,I,J). Based on the Prim's algorithm, list the steps, show the minimum spanning tree (MST) and calculate the resulting total MST length. How many clusters (major and minor) and what are the genes in each cluster? Use Gene A as the initial point. [5 marks]



Question 3

a) Space-filling model is considered to be the most realistic type of representation for protein structures. Give ONE advantage and ONE disadvantage of this model. [2 marks]

b) How do proteins identified from mass spectrometry using online tools? [1 mark]

c) Protein subcellular localization is important for its function. How do we predict the subcellular localization of a protein inside the cell organelles? [2 marks]

d) Supposed you are given a hypothetical protein of unknown function, how might you predict the protein function if the hypothetical protein has homologous proteins? [1 mark]

Continued .....

e) You have a protein sequence, and you want to quickly predict its structure. After performing BLAST and PSI-BLAST searches, you identify the most closely related proteins with known structures as several having 50% amino acid identity to your protein, with a significant expect value. What will you do? Why? [2 marks]

f) Would one expect secondary structure prediction methods intended for soluble proteins to identify membrane helices as helices? Explain. [2 marks]

Question 4

a) What is a RNA chaperone? Discuss the roles of RNA chaperone with the known nature of RNA molecules. [4 marks]

b) What is the "Free Energy Paradigm"? What is the limitation and trend toward the better RNA structure prediction software? [4 marks]

c) Given the matrix below, draw the structure of the RNA. [2 marks]

	C	C	C	U	U	U	U	A	G	G
C	0	0	0	0	0	0	0	1	2	3
C		0	0	0	0	0	0	1	2	3
C			0	0	0	0	0	1	2	2
U				0	0	0	0	1	1	1
U					0	0	0	1	1	1
U						0	0	1	1	1
U							0	1	1	1
A								0	0	0
G									0	0
G										0

END OF PAPER